Enhancing Cancer Registry Data for Comparative Effectiveness Research (CER) Project: Overview and Methodology

Vivien W. Chen, PhD[a], **Christie R. Eheman, PhD**[b], **Christopher J. Johnson, MPH**[c], **Monique N. Hernandez, PhD**[d], **David Rousseau, BS**[e], **Timothy S. Styles, MD, MPH**[b], **Dee W. West, PhD**[f], **Meichin Hsieh, MSPH, CTR**[a], **Anne M. Hakenewerth, PhD**[g], **Maria O. Celaya, MPH, CTR**[h], **Randi K. Rycroft, MSPH, CTR**[i], **Jennifer M. Wike, MBA, MPH**[j], **Melissa Pearson, CTR**[k], **Judy Brockhouse, MPH, CTR**[l], **Linda G. Mulvihill, RHIT,CTR**[b], and **Kevin B. Zhang, PhD**[m]

[a]Louisiana Tumor Registry and Epidemiology Program, School of Public Health, Louisiana State, University Health Sciences Center, New Orleans, Louisiana. [b]Cancer Surveillance Branch, Division of Cancer Prevention and Control, National Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control and Prevention, Atlanta, Georgia. [c]Cancer Data Registry of Idaho, Boise, Idaho. [d]Florida Cancer Data System, Sylvester Comprehensive Cancer Center, University of Miami Miller School of Medicine, Miami, Florida. [e]Rhode Island Cancer Registry, Hospital Association of Rhode Island, Rhode Island. [f]Cancer Registry of Greater California, Public Health Institute, Sacramento, California. [g]Texas Cancer Registry, Cancer Epidemiology and Surveillance Branch, Texas Department of State Health Services, Austin, Texas. [h]New Hampshire State Cancer Registry, Hanover, New Hampshire. [i]Colorado Central Cancer Registry, Colorado Department of Public Health and Environment, Denver, Colorado. [j]CDC-NPCR Contractor, DB Consulting, Silver Springs, MD. [k]North Carolina Cancer Registry, Division of Public Health, Department of Health and Human Services, Raleigh, North Carolina. [l]Alaska Cancer Registry, Anchorage, Alaska. [m]ICF International, Rockville, Maryland.

Abstract

Following the Institute of Medicine's 2009 report on the national priorities for comparative effectiveness research (CER), funding for support of CER became available in 2009 through the American Recovery and Reinvestment Act. The Centers for Disease Control and Prevention (CDC) received funding to enhance the infrastructure of population-based cancer registries and to expand registry data collection to support CER. The CDC established 10 specialized registries within the National Program of Cancer Registries (NPCR) to enhance data collection for all cancers and to address targeted CER questions, including the clinical use and prognostic value of specific biomarkers. The project also included a special focus on detailed first course of treatment for cancers of the breast, colon, and rectum, as well as chronic myeloid leukemia (CML) diagnosed in 2011. This paper describes the methodology and the work conducted by the CDC and the NPCR specialized registries in collecting data for the 4 special focused cancers, including

Address correspondence to Vivien W. Chen, PhD, Louisiana Tumor Registry School of Public Health, Louisiana State University Health Sciences Center, 2020 Gravier Street, 3/F, New Orleans, LA 70112. Fax: (504) 568-5800. vchen@lsuhsc.edu..

the selection of additional data variables, development of data collection tools and software modifications, institutional review board approvals, training, collection of detailed first course of treatment, and quality assurance. It also presents the characteristics of the study population and discusses the strengths and limitations of using population-based cancer registries to support CER as well as the potential future role of population-based cancer registries in assessing the quality of patient care and cancer control.

Keywords

cancer treatment; CER support; methodology; population-based registry

Introduction

In June 2009, the Institute of Medicine (IOM) published a report entitled Initial National Priorities for Comparative Effectiveness Research (CER) and listed 100 CER priorities, including cancer-related objectives.[1] The IOM defined CER as "the generation and synthesis of evidence that compares the benefits and harms of alternative methods to prevent, diagnose, treat, and monitor a clinical condition or to improve the delivery of care. The purpose of CER is to assist consumers, clinicians, purchasers, and policy makers to make informed decisions that will improve health care at both individual and population levels."[1] Funding for CER was provided to the Department of Health and Human Services (HHS) through the American Recovery and Reinvestment Act (ARRA) of 2009, and projects were coordinated through the Agency for Healthcare Research and Quality (AHRQ).

The Centers for Disease Control and Prevention (CDC) received funding to enhance the infrastructure of population-based cancer registries and to support collection of data for CER. CDC developed a project entitled Enhancing Cancer Registry Data for Comparative Effectiveness Research (CER) and in 2010 awarded funds to ICF International and state cancer registries for expanded data collection and 6 special projects. Ten states receiving National Program of Cancer Registries (NPCR) funding that met the eligibility criteria were selected for the CER project based on competitive proposals and were established as specialized cancer registries. These registries had the potential and capability to enhance their infrastructure for "additional data collection, training, methodological development, and expansion of electronic reporting with the goal of supporting comparative effectiveness research and to develop sustainable methods to enhance registry data for public health and research." As part of this CER project, participating NPCR specialized registries expanded data collection to include additional data variables, such as height, weight and smoking status, for all cancers, and they performed linkages with secondary data sets including census data, National Death Index files, hospital discharge data, and the state's breast and cervical cancer early detection programs to enhance registry data. The project also included a special focus on detailed treatment information for cancers of the breast, colon, and rectum as well as chronic myeloid leukemia (CML). An outcome of the project included a dataset that could be used for CER and other research. This paper provides a description of the methodology and the work conducted by the CDC and the NPCR specialized registries in collecting data for the 4 special focused cancers. It also presents the characteristics of the

study population and a discussion of the strengths and limitations of using population-based cancer registries to support CER.

Objectives

The primary objectives of the CER project were to enhance the registry infrastructure and to obtain data that would support CER for cancers of the breast, colon, and rectum as well as CML. The collected data included tumor characteristics (predictive and prognostic biomarkers), stage at diagnosis, first course of treatment (both neoadjuvant and adjuvant), and patient sociodemographic factors and other factors, such as comorbidities and insurance coverage that may influence the choice of treatment options for these patients.

To ensure that current CER needs could be addressed by the project, the CDC in collaboration with AHRQ identified a series of questions, along with the data variables needed to address those and other questions in comparative effectiveness research. The targeted CER issues were:

- Are colon and rectum (colorectal) cancer patients tested for *KRAS* and are the results used appropriately to determine treatment? What impact does *KRAS* testing have on 2-3 year survival among colorectal cancer patients?

- Are rectal cancer patients receiving radiotherapy and what is the timing of radiotherapy? Are disparities apparent in the appropriate neoadjuvant use of radiotherapy among these patients?

- Are CML patients being tested for the BCR-ABL gene and receiving appropriate treatment according to those results?

- Are women with breast cancer being tested appropriately for human epidermal growth factor receptor 2 (HER2), progesterone receptor (PR), and estrogen receptor (ER) status and treated accordingly?

Methods

Case Eligibility

Cases included in the CER project were male and female patients diagnosed in 2011 with either in situ or malignant tumors of the breast or colon/rectum or with CML (see Table 1 for ICD-O-3 site codes and histologies[2]). The 10 participating states included the entire states of Alaska, Colorado, Idaho, Louisiana, New Hampshire, North Carolina, Rhode Island, and Texas, as well as 13 counties of the Sacramento region of California and 5 metropolitan counties of Miami, Florida.

Data Variables

Population-based cancer registries routinely collect North American Association of Central Cancer Registries (NAACCR) standard data variables[3] such as patient demographics, tumor characteristics, cancer stage, treatment, address of residence, and comorbidities. However, these data are insufficient to support contemporary comparative treatment research. Working together with AHRQ, the CDC identified additional data items (Table 2) that were crucial to

address CER questions. The final data variables included in the CER project were: expanded patient information such as height and weight, comorbid conditions, and smoking status; area-based (census tract) socioeconomic status; stage of disease (all Collaborative Stage version 2 [CSv2] data items that are necessary to derive the American Joint Commission on Cancer [AJCC] TNM and Stage for both 6th and 7th editions)[4,5]; tumor biomarkers of prognostic and predictive significance listed under CSv2 site-specific factors (SSFs)[6]; and detailed first course of cancer-directed treatment.[7] For breast cancer, the SSFs included: ER status, PR status, number of positive ipsilateral level I–II axillary lymph nodes, presence of isolated tumor cells in regional lymph nodes, size of invasive component of the tumor, Nottingham or Bloom-Richardson tumor grade/score, HER2, response to neoadjuvant therapy, and multigene signature testing. For colorectal cancer, the SSFs collected were: carcinoembryonic antigen (CEA), clinical assessment of regional lymph nodes, presence of tumor deposits, tumor regression grade, circumferential resection margin (CRM), microsatellite instability (MSI), perineural invasion, *KRAS* testing, and loss of heterozygosity (LOH). As for CML, Janus Kinase 2 (JAK2) gene mutation was collected in addition to BCR-ABL testing.

In order to evaluate comparative effectiveness of treatments for cancers of the breast, colon, rectum, and CML, complete and accurate first course of cancer-directed treatment was essential. First course of cancer treatment was defined as the therapy regimen that was given or planned at the time of initial diagnosis, prior to disease recurrence or progression.[7] While central cancer registries routinely collect detailed information on surgery (which often is performed in hospital at inpatient or outpatient settings) and radiation (hospital-affiliated and freestanding), other adjuvant treatments have often been missing or incomplete. Therefore, the CER project focused on collecting complete adjuvant treatment occurring within 12 months of diagnosis, particularly detailed chemotherapy data. To identify each drug consistently across the registries, abstractors used the Cancer Chemotherapy National Service Center (NSC) number. The NSC is a numeric identifier for substances submitted to the National Cancer Institute (NCI) for testing and evaluation during its investigational phase and a registration number for the Developmental Therapeutics Program (DTP) repository. Abstractors obtained NSC numbers from the Web-based version of the SEER*Rx Interactive Antineoplastic Drugs Database[8] that provides detailed drug information including generic and brand name, abbreviation, NSC number, drug category and subcategory and is readily available on the NCI website (**http://seer.cancer.gov/ seertools/seerrx**).

The CER chemotherapy variables included each chemotherapy agent's name and NSC number, number of chemotherapy cycles planned, total dose planned, number of doses received, and total dose received, start and end dates of chemotherapy, and whether chemotherapy was completed as planned, as well as growth factor agents (granulocyte, erythrocyte, and thrombocyte) that might have been given. Patient's height and weight were also collected, especially for CER patients receiving chemotherapy. In addition, hormonal therapy and biological response modifier use were documented. If any subsequent treatment was given within 12 months of diagnosis as a result of recurrence, all subsequent treatment modalities were also recorded when available.

A CDC work group developed a data dictionary for the non-NAACCR standard variables, with an emphasis on capturing all cancer therapies received within 12 months of diagnosis. An oncologist provided consultation on the selection and definitions of additional variables of CER significance. Once a variable was identified as necessary and available in medical records, it was added to the data dictionary with a description, rationale, codes, and coding instructions. The work group used established, recognized guidelines for cancer data collection whenever possible. As an example, the SEER*Rx database was used to assign the NSC number for systemic therapy because it provides a consistent method to identify each drug. A few drugs that were either not listed in SEER*Rx or did not have an assigned NSC number were assigned an identifying number to be used by all registries.

Data Collection Tool and Software

The participating states use different registry software for routine data collection and have various database management systems. There was no existing electronic abstraction tool for the CER project because the non-NAACCR standard data variables had not previously been collected. The CER data dictionary was provided to the registries' software vendors for modification of their software so that these data items could be abstracted and be incorporated into the registry database. Different approaches were used by the specialized registries based on the capability of their software vendors, experience of their information technology (IT) staff, reporting requirement notification process, and CDC technical support. As a result, each registry had its own CER-specific data abstracting tools, edit programs and consolidation logic. However, all registries used a standard edits metafile and the same record layout for data submission.

To ensure data quality and consistency CDC staff developed a set of single field and inter-field edits that all CER participating registries ran prior to data submission.

Institutional Review Board Approvals

The collection of additional data items for the CER project was authorized by existing cancer reporting laws and regulations in each state. No patients were contacted for the project and the collection of enhanced registry data was considered by most states to be public health surveillance and, as such, exempt from institutional review board (IRB) approval requirements. One state amended its cancer registry reporting regulations to add the nonstandard data items to its surveillance reporting requirements. Because the CER project title included the word "research," some cancer treatment facilities required further explanation or, in some cases, sought full IRB review.

It should be noted that while the project activities described herein did not include research, but focused on collection of the data needed to support CER, all subsequent data analyses, conducted after the completion of project, must be covered under applicable IRB requirements of the investigators.

Training

Training was conducted by the CDC and its contractors before data collection began. The objectives of the CER project were explained, and the data variables and data dictionary,

especially those items not routinely collected by the registries, were reviewed and discussed. Trainers provided instructions on data collection and coding rules, including CSv2 stage; SSFs, especially biomarkers such as *KRAS* and HER2; complete first course of treatment which included surgery, chemotherapy, hormonal therapy, biological response modifiers (including bone marrow transplant, stem cell harvest, surgical or radiation endocrine therapy); growth factors; and BCR-ABL tests for CML.

In addition to CDC central training, all participating specialized registries offered subsequent in-house training and ongoing education, depending on the needs and experience of their abstractors. Various training modes included workshops, webinars, webcasts, teleconferences, and presentations at the state cancer registrars' associations meetings. All states conducted one-on-one, in-person and/or webcast training when needed, and some also offered onsite demonstration. The intensity and extent of training varied by registry and depended on their available resources. For example, one registry housed at the health sciences center received training, from a clinical oncology nurse, on the National Comprehensive Cancer Network (NCCN) treatment guidelines for each cancer by tumor size, nodal involvement, and distant metastasis, as well as by biomarker/prediction of treatment response (eg, HER2, ER/PR and *KRAS*) and menopausal status (breast only).[9] This training detailed first course of treatment, sequence of treatment modalities, common chemotherapy regimens, chemotherapy drug names, "standard" cycle and dose, and calculation of total dose (planned and received). In addition, the training included explanation and discussion on the chemotherapy "flow chart," common side effects, and toxicities. Another registry created an abstracting workbook for data collection which provided information not contained in the data dictionary, such as treatment guidelines based on stage of disease and tips for interpreting laboratory reports for biomarkers. Other registries used highly experienced certified tumor registrars (CTRs), or combined CTR and registered nurse, to conduct trainings with hospital-based and non-hospital-based CER data collection staff. One registry had its abstractors practice CER data collection on cases diagnosed prior to 2011, interacted with hospital and nonhospital staff, and established remote access to some facilities. All participating registries conducted refresher trainings, after data collection was in progress, to ensure accurate and consistent interpretation of the information and coding of the data. Every registry designated a contact staff person to address CER data collection issues, answer questions, and provide clarifications.

The CDC data dictionary work group reviewed all training materials developed by the states to ensure accuracy and consistency. Many states shared and customized training materials which were made available through the CDC CER Information-Sharing Portal.

Data Collection

As noted previously, all NPCR specialized registries have legislative rules on cancer reporting which provide authorization to access medical records and collect cancer-related data variables from hospitals and nonhospital settings. Because of the vast number of cancer cases (more than a quarter of the annual caseload) and the large number of additional data variables collected, the majority of specialized registries hired and trained additional

abstractors for this project; many reassigned existing staff and some used a combination of new and existing staff.

a. Hospital—When CER data collection started in 2012, a majority of the 2011 cases were already reported to the central registries. Building on the routine NAACCR abstracts they had on the CER cases, the hospital registrars (or central registry abstractors who visited the hospitals where the cancer was initially diagnosed) collected as many of the additional CER data items as possible. Most adjuvant treatment was given after the patient was discharged from the hospital; it was therefore necessary for abstractors to identify the facilities and/or physicians who followed the patients and provided adjuvant treatments. Some registries requested that hospital abstractors delay their submission of CER cancer cases until after the first course of treatment was complete and all information that could be obtained was abstracted. However, most CER cases required data collection from all source documents, including both hospital and nonhospital settings. On very rare occasions, where hospitals used a unified chart system that provided inpatient and outpatient medical records in one single database, abstractors were able to record all therapy modalities from a single setting.

b. Nonhospital sources—Adjuvant radiation was generally obtained from hospital outpatient or freestanding radiation centers, while information on chemotherapy was obtained mostly from hospital-based or independent hematology/oncology practice groups. Most specialized registries hired additional data specialists to collect CER data at physician offices and treatment facilities. This included identifying noncancer registry reporting sources to be targeted for data collection, developing methods of data collection and coordinating with the reporting source to carry out case identification and data collection.

Registries used 3 primary methods of data collection: onsite visits for abstracting, which was the most common method, obtaining hard copies or securing remote access to the facility's electronic medical record, and, on rare occasions, transmission of data by the facility. Abstraction at nonhospital sources usually required an initial onsite visit to learn how to navigate each practice's electronic or paper medical record systems, and follow-up visits to capture missing information. Follow-up was also done by phone, fax, or email whenever possible. Multiple visits to more than one physician office were frequently necessary to complete the first course of chemotherapy and other adjuvant treatment. Ascertaining the dose for each cycle of every chemotherapy agent was very challenging, and information was not always available, making the calculation of total chemotherapy dose received extremely difficult, often necessitating multiple visits.

Once data were obtained from all sources, CER staff reviewed and edited the cases for accuracy, consistency, and completeness. Cases were followed back to treating physician and/or facility to obtain missing information. First course of treatment received within 12 months of diagnosis was edited and consolidated so that the data could be provided for comparative effectiveness of treatments. Patient address of residence at the time of diagnosis was geocoded to census tract by each registry and used to link case data with Census Bureau data files for census tract level socioeconomic indicators, such as poverty level, employment status, and urbanization.[10]

Data Quality

As questions or issues related to data collection occurred, each state submitted the inquiry to a technical assistance folder on a central CER-specific Web portal. The dedicated Web portal was maintained by the CER contractor and used by all participants for project-related information sharing and documentation. States could post questions and receive timely responses that could be viewed and systematically searched by all collaborators. By managing the technical assistance requests in this manner, all states had access to the same information in real time and previous issues could be reviewed as often and by as many CER team members as needed. The responses could also be exported to a spreadsheet format for use by analysts and other members. As issues arose, a CDC data items work group reviewed each question, drafted a consensus response, and shared with oncologists and contractors, as appropriate. Once the response was final, it was posted to ensure standardization and consistency across all states.

All registries conducted periodic review of abstracted cases throughout the study period, using source documents to ensure high quality data and consistent interpretation of medical record information. When any systemic problem was identified, the registry immediately issued an alert to all data collectors. If necessary, training or refresher courses were offered. Although most states did not perform reabstraction audits, all performed reabstraction for cases with missing, unknown, or questionable data.

All specialized registries also ran their data through the NAACCR Hispanic Identification Algorithm (NHIA)[11] and the NAACCR Asian/Pacific Islander Identification Algorithm (NAPIIA)[12] as well as participated in linkages with the Indian Health Service to improve the quality of their data on race and ethnicity.[13]

Results

Among the 10 participating states, a total of 75,042 cancer patients were diagnosed in 2011 with breast, colon, or rectum cancer, or CML. Breast cancer accounted for 64.6% of all cases, about one-quarter (24.5 %) of patients were diagnosed with colon cancer, 9.5% with rectal cancer, and CML accounted for 1.4%. The demographic and tumor characteristics of these patients are shown in Table 3. The overall high percentage of female patients (80.6%) reflects the large number of breast cancer cases that account for almost two-thirds of the total CER cases. There was no difference in gender distribution for colon cancer, and a higher percentage of rectal cancer and CML were males. Approximately half (48.2%) of the colon cancer cases were diagnosed in patients 70 years and older whereas for breast cancer, only about a quarter (27.8%) were aged 70 or older. The majority of patients were non-Hispanic whites (68.2%), non-Hispanic blacks, and Hispanics (of all races) each represented about 14%, and less than 4% were of other or unknown races. The racial/ethnic pattern was very similar for all 4 cancer groups.

Overall, approximately half of the patients had private insurance (including Medicare with private supplement or Medicare Advantage Plans), 29% had Medicare only, 10% had Medicaid, 2.5% had other public insurance such as Tricare, Veterans' Affairs, or Indian Health Services, and less than 5% had no insurance. Breast cancer patients (54%) were most

likely to have private insurance whereas colon cancer patients were the least likely (38%). The large proportion of Medicare coverage among colon cancer patients (37%) reflects the older age at diagnosis compared with other cancers. Over half (54.5%) of all patients had at least 1 comorbidity, with the comorbid conditions being more prevalent in colorectal than breast cancer patients. The majority of CER patients lived in census tracts designated as urban or urban/rural mixed, and less than 10% lived in rural area. About 83% resided in census tracts with less than 20% of families having income below federal poverty guidelines. Information on cigarette smoking was missing in 34.5% of the patients. Among those with known information, 62% never smoked cigarettes, 24% were former users, and 14% were current smokers.

There was considerable variation of tumor stage at diagnosis by cancer type. About two-thirds (68%) of breast cancer patients were diagnosed with early stage disease (in situ and localized) based on the SEER Summary Stage 2000, contrasting with only 40% among colorectal cancer patients. Similar tumor stage pattern was observed for AJCC TNM Stage Group.

The distribution of cancer cases also varied greatly by state (ranging from 819 to 26,752), reflecting the population size of the participating states and geographic areas.

Discussion

When the National Cancer Policy Board of the IOM, National Academy of Sciences issued a report on *Ensuring Quality of Cancer Care* in 1999, it concluded that "....for many Americans with cancers, there is a wide gulf between what could be construed as the ideal and the reality of their experience with cancer care."[14] Its subsequent report of Enhancing Data Systems to Improve the Quality of Cancer Care[15] in 2000 recommended the need for comprehensive cancer data systems that could be used to gauge the status of cancer care and measure quality. It identified 3 existing national cancer surveillance programs that could be used for quality improvement of cancer care. They were: the NPCR of CDC; the Surveillance, Epidemiology and End Results (SEER) program of the NCI; and the National Cancer Data Base (NCDB), sponsored by the American College of Surgeons' Commission on Cancer (ACoS-CoC) and the American Cancer Society. While each of these data systems has its own limitations, they also each hold great potential and could be enhanced to assess and improve quality of cancer care in the nation.

In response to the IOM recommendations, CDC-NPCR initiated a pattern of care study, comparing the observed patterns of care with the accepted guidelines for localized breast cancer, localized prostate cancer and Stage III colon cancer, in conjunction with the international CONCORD Study.[16-18] Subsequently the CDC, in collaboration with researchers from 7 central cancer registries, conducted the Breast and Prostate Cancer Data Quality and Patterns of Care (POC-BP) Study in 2007-2009.[19] The study examined first course of cancer treatment received and how the patterns of care varied by patient, provider and other health system level factors. It also evaluated whether the care was concordant with nationally recognized treatment guidelines, given the presence and severity of comorbidities.[20-24]

The NCI has also conducted patterns of care/quality of care (POC) studies since 1987 under a congressional mandate [Public Law 100-607, Sec. 413 (a)(2)(C) adopted November 4, 1988]. The collection of NCI POC data is coordinated jointly by the Division of Cancer Control and Population Sciences and the Division of Cancer Treatment and Diagnosis. Using the infrastructure of SEER registries, population-based samples of cases are selected every year for various cancer types to evaluate the dissemination of cancer therapy or guideline care into community practice and to identify possible determinants of dissemination and variations in therapy.[25] Findings from POC studies provide educational or training opportunities for professional societies and public health groups to improve the quality of cancer care and reduce disparities in treatment and survival among different population groups. A wide range of cancer care has been studied by NCI over the years, including the trends in using adjuvant multi-agent chemotherapy and tamoxifen for breast cancer;[26] age, sex, and racial differences in the use of standard adjuvant therapy for colorectal cancer;[27] and clinical trial participation and time to treatment among adolescents and young adults with cancer.[28] Though it is not population-based, researchers have also used the NCDB to assess treatment patterns and their determinants among patients treated in facilities participating in the CoC-accredited cancer programs. Examples include assessing the importance of socioeconomic status and treatment institution in neo-adjuvant therapy for Stage IIIA non-small cell lung cancer[29] and the impact of facility volume on therapy and survival for advanced cervical cancer.[30]

In 2009, CDC received funding to establish 10 specialized cancer registries to collect additional data to support CER. This CER project is the largest and most comprehensive data collection effort ever conducted by population-based cancer registries in the United States to obtain complete and detailed first course of cancer-directed treatment for breast, colon, and rectal cancers and CML. This project covers more than a quarter (27.3%) of the US population, including a very high representation among minority populations. In fact, the combined catchment areas of the CER project include approximately 25% of African Americans, 37% of Asian/Pacific Islanders, 32% of American Indians/Alaska natives, and 44% of Hispanics living in the United States.[31] This provides a unique opportunity to examine the patterns of care for these cancers in a racially and ethnically diverse population, to compare the benefits and risks of alternative treatments and care deliveries as well as to assess their outcomes. The project includes all newly-diagnosed cancer patients residing in the participating states in care settings that are representative of contemporary practice in the 10 geographic areas across the United States. Previous projects of this nature were often conducted in single major medical centers or in clinical trial groups where the patients were not representative of all US patients, and the treatments were not representative of contemporary practice of both major urban cancer centers and small rural hospitals.[32-34]

Using population-based central cancer registries to collect data for CER has numerous strengths. Existing registry infrastructure and authorizing laws and regulations make it feasible to expedite the process, implement the procedures and collect additional cancer data in a timely manner. This registry infrastructure provided a baseline of well-established standard definitions and codes for cancer reporting, tumor staging and treatment/drugs. Because central cancer registries have been in place for several decades, registry staff are familiar with cancer reportability and the standard rules and codes. They have extensive

experience with medical records abstraction and are knowledgeable of various cancer treatment modalities, requiring fewer staff and less training. Over time, central cancer registries have built excellent rapport with cancer care facilities and providers, allowing them to gain access to the facilities and/or connect remotely to the electronic medical records, resulting in an efficient and cost-effective means of obtaining the additional data for CER. In addition, population-based registries include all newly-diagnosed cancer patients; therefore findings using their data or a representative sample can be generalized to the US general population. There are also, however, limitations. Hospitals are required to submit cancer cases to central registries within 6 months after diagnosis, when first course of cancer treatment is often still ongoing; thus only partial treatment information is recorded at the time of initial report. While some states ask hospitals to submit update records which contain the additional treatment information, most state registries do not include these update records due to limited staff and competing registry priorities. Information on adjuvant therapies, especially hormonal and chemotherapy that are provided in nonhospital settings, is sometimes missing or incomplete. Collecting additional data that are not standard registry variables but are relevant to address CER topics requires special trainings. Furthermore, efforts and resources spent on project-specific activities such as software modification, training, data edits and added data submissions compete with regular registry operations and functions, potentially causing delays in routine registry data collection.

Despite the limitations, the CDC and the 10 NPCR specialized cancer registries have demonstrated the feasibility of utilizing population-based cancer registries to support comparative effectiveness investigations and other research in a cost-effective and timely manner. Moving forward, the CDC and the specialized cancer registries will continue to expand the uses of registry data beyond measuring cancer burden and evaluating stage shift by examining diagnostic procedures and cancer treatments and assessing how different procedures and treatment modalities impact patient outcomes of recurrence and survival. Central cancer registries provide a unique population base for cancer research. In addition, the consolidation of data from multiple locations of care is unique to the central cancer registry. No other disease surveillance system in the United States is as standardized and comprehensive as cancer surveillance, allowing researchers to track the outcomes of changes in the medical care system and assess the effectiveness of public health interventions. The CDC's specialized cancer registries and the entire cancer registration program are well positioned to conduct comparative effectiveness research, evaluate patient outcome, and inform future medical practice.

Acknowledgments

References

1. Institute of Medicine (US) Committee on Comparative Effectiveness Research Prioritization, Board on Health Care Services. Initial National Priorities for Comparative Effectiveness Research. National Academies Press; Washington, DC: 2009.

2. Fritz, AG. International Classification of Diseases for Oncology: ICD-O. 3rd ed.. World Health Organization; Geneva, Switzerland: 2000.

3. North American Association of Central Cancer Registries. Record Layout Version 12.1. 15th ed.. North American Association of Central Cancer Registries; Springfield, IL: 2010. Standards for Cancer Registries, Volume II: Data Standards and Data Dictionary..

4. Greene, FL.; American Joint Committee on Cancer, American Cancer Society. AJCC Cancer Staging Handbook: From the AJCC Cancer Staging Manual. 6th ed.. Springer; New York, NY: 2002.

5. Edge, SB.; American Joint Committee on Cancer, American Cancer Society. AJCC Cancer Staging Handbook: Fiom the AJCC Cancel Staging Manual. 7th ed.. Springer; New York, NY: 2010.

6. American Joint Committee on Cancer. [April 30, 2013] Collaborative Stage Version 2. http://web2.facs.org/cstage0204/schemalist.html.

7. Commission on Cancer. Facility Oncology Registry Data Standards Revised for 2011. American College of Surgeons; Chicago, IL: 2011. p. 199-284.http://www.facs.org/cancer/coc/fords/FORDS_for_2011 _01012011.pdf. [January 15, 2012]

8. SEER. RX interactive antineoplastic drug database [database online]. http://seer.cancer.gov/seertools/seerrx

9. National Comprehensive Cancer Network. [January 15, 2012] NCCN guidelines for treatment of cancer. http://www.nccn.org/professionals/physician_gls/f_guidelines.asp.

10. Krieger N, William DR, Moss NE. Measuring social class in US public health research: concept, methodologies and guidelines. Annu Rev Public Health. 1997; 18:341–378. [PubMed: 9143723]

11. NAACCR Race and Ethnicity Work Group. NAACCR Guideline for Enhancing Hispanic/Latino Identification: Revised NAACCR Hispanic/ Latino Identification Algorithm [NHIA V2.2.IJ. North American Association of Central Cancer Registries; Springfield, IL: 2011.

12. NAACCR Race and Ethnicity Work Group. NAACCR Asian Pacific Islander Identification Algoiithm [NAPIIA vl.2.1]. North American Association of Central Cancer Registries; Springfield, IL: 2011.

13. Espey DK, Wiggins CL, Jim MA, Miller BA, Johnson CJ, Becker TM. Methods for improving cancer surveillance data in American Indian and Alaska Native populations. Cancel. 2008; 3(5 Suppl):11, 1120–1130. doi: 10.1002/cncr.23724.

14. Hewitt, M.; Simone, JV., editors. Ensuring Quality Cancer Care. National Academy Press; Washington, DC: 1999.

15. Hewitt, M.; Simone, JV., editors. Enhancing Data Systems to Improve the Quality of Cancer Care. National Academy Press; Washington, DC: 2000.

16. McDavid K, Schymura MJ, Armstrong L, et al. Rationale and design of the National Program of Cancer Registries' breast, colon, and prostate cancer patterns of care study. Cancer Causes Control. 2004; 15(10):1057–1066. [PubMed: 15801489]

17. Alley LG, Chen VW, Wike JM, et al. CDC-NPCR's breast, colon, and prostate cancer data quality and patterns of care study: overview and methodology. J Registry Manage. 2007; 34:148–157.

18. Alley LG, Fulton JP, Wike JM, et al. Studying patterns of care: an evaluation of a project using CDC-NPCR data. J Registry Manage. 2008; 35(1):27.

19. German RR, Wike JM, Bauer KR, et al. Quality of cancer registry data: findings from CDC-NPCR's breast and prostate cancer data quality and patterns of care study. J Registry Manage. 2011; 38(2):75–86.

20. Fleming ST, Hamilton AS, Sabatino SA, et al. Treatment patterns for prostate cancer: comparison of medicare claims data to medical record review. Med Care. 2014; 52(9):e58–64. doi: 10.1097/MLR.0b013e318277eba5. [PubMed: 23222532]

21. Hamilton AS, Wu XC, Lipscomb J, et al. Regional, provider, and economic factors associated with the choice of active surveillance in the treatment of men with localized prostate cancer. J Natl Cancer Inst Monogr. 2012; 21(45):3–220.

22. Fleming ST, Kimmick GG, Sabatino SA, et al. Defining care provided for breast cancer based on medical record review or medicare claims: information from the centers for disease control and prevention patterns of care study. Ann Epidemiol. 2012; 22(11):807–81 3. [PubMed: 22948184]

23. Wu XC, Lund MJ, Kimmick GG, et al. Influence of race, insurance, socioeconomic status, and hospital type on receipt of guideline-concordant adjuvant systemic therapy for locoregional breast cancers. J Clin Oncol. 2012; 30(2):142–150. [PubMed: 22147735]

24. Fleming ST, Sabatino SA, Kimmick G, et al. Developing a claim-based version of the ACE-27 comorbidity index: a comparison with medical record review. Med Care. 2011; 49(8):752–760. [PubMed: 21490514]

25. Patterns of care/quality of care. National Cancer Institute website; http://appliedresearch.cancer.gov/surveys/poc/. [September 13, 2014]

26. Mariotto A, Feuer EJ, Harlan LC, Wun L- M, Johnson K, Abrams J. Trends in use of adjuvant multi-agent chemotherapy and tamoxifen for breast cancer in the United States: 1975-1997. J Natl Cancer Inst. 2002; 94:1626–1634. [PubMed: 12419789]

27. Potosky AL, Harlan LC, Kaplan RS, Johnson KA, Lynch CF. Age, sex, and racial differences in the use of standard adjuvant therapy for colorectal cancer. J Clin Oncol. 2002; 20:1192–1120. [PubMed: 11870160]

28. Parsons HM, Harlan LC, Seibel NL, Stevens JL, Keegan TH. Clinical trial participation and time to treatment among adolescents and young adults with cancer: does age at diagnosis or insurance make a difference? J Clin Oncol. 2011; 29(30):4045–4053. [PubMed: 21931022]

29. Sher DJ, Liptay MJ, Fidler MJ. Prevalence and predictors of neoadjuvant therapy for stage MIA non-small cell lung cancer in the National Cancer Database: Importance of socioeconomic status and treating institution. Int J Radiat Oncol Biol Phys. 2014; 89(2):303–312. [PubMed: 24685443]

30. Lin JF, Berger JL, Krivak TC, et al. Impact of facility volume on therapy and survival for locally advanced cervical cancer. Gynecol Oncol. 2014; 132(2):416–22. [PubMed: 24333361]

31. US Census Bureau's population estimates program. [February] National Cancer Institute website. 2014. http://seer.cancer.gov/popdata.thru.2011/download.html.

32. Erickson BK, Doo DW, Zhang B, Huh WK. Black race independently predicts worse survival in uterine carcinosarcoma. Cynecol Oncol. 2014; 133(2):238–41.

33. Lee S, Reha JL, Tzeng CWD, Massarweh NN. Race does not impact pancreatic cancer treatment and survival in an equal access federal health care system. Ann Surg Oncol. 2013; 20:4073–4079. [PubMed: 24002535]

34. Unger JM, Barlow WE, Martin DP, et al. Comparison of survival outcomes among cancer patients treated in and out of clinical trials. J Natl Cancer Inst. 2014; 106(3):dju002. [PubMed: 24627276]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

Eligibility Criteria for Comparative Effectiveness Research Cases

Site	ICD-O-3 Site Code	Histology	Behavior	Sex	Diagnosis Year
Breast	C50.0–C50.9	All except 9050–9055, 9140, and 9590–9992	In situ, malignant	Male and female.	2011
Colon	C18.0–18.9 [*]	All except 9050–9055, 9140, and 9590–9992	In situ, malignant	Male and female	2011
Rectum	C19.9, C20.9	All except 9050–9055, 9140, and 9590–9992	In situ, malignant	Male and female	2011
Chronic Myeloid Leukemia	C42.1	Include 9863, 9875, 9876, 9945, and 9946	malignant	Male and female	2011

[*] Including Appendix (site code = C18.1).

Table 2

Non-NAACCR Standard Data Variables[*] Defined and Collected for the Comparative Effectiveness Research Project

Patient characteristics
• Height
• Weight
• Comorbidities: up to 10 standard NAACCR comorbidities
• Tobacco Use: cigarette, smoking tobacco products other than cigarettes (eg, pipes, cigars, kreteks), smokeless tobacco, not otherwise specified
• Socio-Economic Status Indicators
A wide range of indicators linked to cancer incidence data from multiple US Census Bureau data files, including area level indicators for:
• Urban/rural
• Poverty level
• Health insurance coverage
• Income ranges
• Employment status
• Occupation and Industry data
• Class of worker (blue collar, white collar)
Diagnostic work up for CML
• BCR-ABL Cytogenetic Result and Date
• BCR-ABL FISH Result and Date
• BCR-ABL Qualitative Result and Date
• BCR-ABL Quantitative and Date
First course treatment information:
• Chemotherapy: agent and NSC number (for up to 6 chemotherapy agents)
• Chemotherapy: number of doses planned and received (for up to 6 chemotherapy agents)
• Chemotherapy: dose amount and units planned and received (for up to 6 chemotherapy agents)
• Chemotherapy: administration start and end dates (for up to 6 chemotherapy agents)
• Chemotherapy: completion status
• Chemotherapy: Granulocyte CSF Status
• Chemotherapy: Erythrocyte Growth Factor Status and Thrombocyte Growth Factor Status
• Hormone: agent and NSC number (for up to two hormone agents)
• Biologic Response Modifier (BRM): agent and NSC number (for up to 2 hormone agents)
Subsequent/second course treatment information:
• Reason Subsequent Treatment
• Subsequent treatment date
• Subsequent Surgery
• Subsequent Radiation
• Subsequent Chemotherapy and Chemo NSC (for up to 6 chemo agents)
• Subsequent Hormone and hormone NSC (for up to 2 hormone agents)
• Subsequent BRM and BRM NSC (for up to 2 BRM agents)
• Subsequent Transplant/Endocrine
• Subsequent Other

*Only variables defined and collected exclusively for Comparative Effectiveness Research (CER) activities are included above. There are many other variables included in the CER data collection activities that are routinely collected by National Program of Cancer Registries (NPCR) and defined by North American Association of Central Cancer Registries' (NAACCR's) *Standards for Cancer Registries, Volume II: Data Standards and Data Dictionary, Fifteenth Edition, Record Layout Version 12.1.*

Table 3

Demographic and Tumor Characteristics of Comparative Effectiveness Research Patients by Cancer Site, 2011

	Breast		Colon[1]		Rectum		CML[2]		Total	
	N	%	N	%	N	%	N	%	N	%
Total	48,456	64.6%	18,413	24.5%	7,129	9.5%	1,044	1.4%	75,042	100.0%
Sex										
Male	372	0.8%	9,361	50.8%	4,186	58.7%	609	58.3%	14,528	19.4%
Female	48,079	99.2%	9,044	49.1%	2,942	41.3%	435	41.7%	60,500	80.6%
Other (hermaphrodite)	3	0.0%	1	0.0%	0	0.0%	0	0.0%	4	0.0%
Transsexual	2	0.0%	1	0.0%	1	0.0%	0	0.0%	4	0.0%
Not State/Unknown	0	0.0%	6	0.0%	0	0.0%	0	0.0%	6	0.0%
Age at Diagnosis (years)										
<50	10,068	20.8%	1,709	9.3%	1,049	14.7%	273	26.1%	13,099	17.5%
50-59	11,883	24.5%	3,226	17.5%	1,849	25.9%	160	15.3%	17,118	22.8%
60-69	13,024	26.9%	4,597	25.0%	1,844	25.9%	197	18.9%	19,662	26.2%
70	13,479	27.8%	8,881	48.2%	2,387	33.5%	414	39.7%	25,161	33.5%
Unknown	1	0.0%	0	0.0%	0	0.0%	0	0.0%	1	0.0%
Race/Ethnicity										
Non-Hispanic White	33,502	68.8%	12,263	66.6%	4,722	66.1%	697	67.1%	51,184	68.2%
Non-Hispanic Black	6,456	13.3%	2,777	14.9%	947	13.0%	132	13.2%	10,312	13.7%
Non-Hispanic Other[3]	1,183	2.3%	373	1.9%	187	2.5%	18	1.5%	1,761	2.3%
Hispanic (all races)	6,679	13.3%	2,772	14.7%	1,166	16.2%	180	15.7%	10,797	14.4%
Oth/Unk/Missing	636	2.3%	228	2.0%	107	2.2%	17	2.4%	988	1.3%
Health Insurance										
No Insurance	1,820	3.8%	1,011	5.5%	470	6.6%	80	7.7%	3,381	4.5%
Private Insurance	26,335	54.3%	6,982	37.9%	3,173	44.5%	411	39.4%	36,901	49.2%
Public: Medicaid	4,620	9.5%	1,865	10.1%	770	10.8%	112	10.7%	7,367	9.8%
Public: Medicare	12,228	25.2%	6,793	36.9%	2,031	28.5%	322	30.8%	21,374	28.5%
Public: Other[4]	901	1.9%	640	3.5%	290	4.1%	18	1.7%	1,849	2.5%

	Breast		Colon[1]		Rectum		CML[2]		Total	
	N	%	N	%	N	%	N	%	N	%
Unknown	2,446	5.0%	1,078	5.9%	379	5.3%	100	9.6%	4,003	5.3%
Missing/Blank	106	0.2%	44	0.2%	16	0.2%	1	0.1%	167	0.2%
Comorbidity Data										
None documented	23,796	49.1%	6,540	35.5%	2,754	38.6%	459	44.0%	33,549	44.7%
At least one	24,233	50.0%	11,737	63.7%	4,335	60.8%	579	55.5%	40,884	54.5%
Blank	427	0.9%	136	0.7%	40	0.6%	6	0.6%	609	0.8%
Census Tract Residence[5]										
Urban	28,060	57.9%	10,358	56.3%	3,935	55.2%	586	56.1%	42,939	57.2%
Rural	3,871	8.0%	1,653	9.0%	665	9.3%	77	7.4%	6,266	8.3%
Mixed	15,746	32.5%	5,988	32.5%	2,385	33.5%	359	34.4%	24,478	32.6%
Missing	779	1.6%	414	2.2%	144	2.0%	22	2.1%	1,359	1.8%
Census Tract Poverty[6]										
Not in poverty	40,920	84.4%	14,785	80.3%	5,721	80.2%	839	80.4%	62,265	83.0%
Poverty	6,744	13.9%	3,203	17.4%	1,262	17.7%	181	17.3%	11,389	15.2%
Blank/Missing	792	1.6%	426	2.3%	146	2.0%	24	2.3%	1,388	1.8%
Cigarette Use										
Never Used	20,692	42.7%	7,047	38.3%	2,489	34.9%	408	39.1%	30,636	40.8%
Current user	4,078	8.4%	1,696	9.2%	927	13.0%	84	8.0%	6,785	9.0%
Former user	7,092	14.6%	3,141	17.1%	1,311	18.4%	197	18.9%	11,741	15.6%
Unknown	16,594	34.2%	6,529	35.5%	2,402	33.7%	355	34.0%	25,880	34.5%
SEER Summ Stg 2000[7]										
In Situ	8,796	18.2%	824	4.5%	288	4.07a			9,908	13.2%
Localized	24,090	49.7%	6,289	34.2%	2,649	37.2%			33,028	44.0%
Regional	11,312	23.3%	5,991	32.5%	2,225	31.2%			19,528	26.0%
Distant	2,368	4.9%	3,731	20.3%	1,256	17.6%	1,044	100.0%	8,399	11.2%
Unknown	1,890	3.9%	1,578	8.6%	711	10.0%			4,179	5.6%
Derive AJCC Stg 7th Ed[8]										

	Breast		Colon[1]		Rectum		CML[2]		Total	
	N	%	N	%	N	%	N	%	N	%
Stage 0	8,804	18.2%	1,132	6.1%	443	6.2%			10,379	13.8%
Stage I	18,037	37.2%	3,514	19.1%	1,717	24.1%			23,268	31.0%
Stage II	11,714	24.2%	4,218	22.9%	1,240	17.4%			17,172	22.9%
Stage III	4,215	8.7%	4,122	22.4%	1,587	22.3%			9,924	13.2%
Stage IV	2,291	4.7%	3,568	19.4%	1,170	16.4%			7,029	9.4%
NA/Unknown	3,395	7.0%	1,859	10.1%	972	13.6%	1,044	100.0%	7,270	9.7%
State of Residence										
AK	530	1.1%	214	1.2%	68	1.0%	7	0.7%	819	1.1%
CA-Sacramento[9]	3,328	6.9%	1,104	6.0%	440	6.2%	47	4.5%	4,919	6.6%
CO	4,198	8.7%	1,354	7.4%	520	7.3%	83	8.0%	6,155	8.2%
FL–Metro Miami[10]	6,734	13.9%	2,976	16.2%	1,019	14.3%	121	11.6%	10,850	14.5%
ID	1,231	2.5%	468	2.5%	187	2.6%	28	2.7%	1,914	2.6%
LA	3,915	8.1%	1,721	9.3%	705	9.9%	100	9.6%	6,441	8.6%
NC	9,177	18.9%	2,970	16.1%	1,165	16.3%	190	18.2%	13,502	18.0%
NH	1,454	3.0%	464	2.5%	193	2.7%	18	1.7%	2,129	2.8%
RI	1,057	2.2%	340	1.8%	151	2.1%	13	1.2%	1,561	2.1%
TX	16,832	34.7%	6,802	36.9%	2,681	37.6%	437	41.9%	26,752	35.6%

[1] Appendix cases are included from Colon.

[2] CML = Chronic Myeloid Leukemia.

[3] Non-Hispanic Other includes American Indian/Alaskan Native /Asian or Pacific Islander.

[4] Public: Other includes Tricare, Military, VA, Indian Health Service (IHS).

[5] A census tract residence was considered "Urban" if all households in that census tract were considered to be in an urban setting as defined by the Census Bureau; A census tract residence was considered "Rural" if all households in that census tract were considered to be in a rural setting as defined by the Census Bureau; A census tract residence was considered "Mixed" if some of the households in the census tract were considered to be in an urban setting and some in a rural setting as defined by the Census Bureau.

[6] Classification of Census Tract Poverty based on Krieger et al (ref 10). Not in poverty defined as: <20% of census tract families had income below poverty line in last 12 months; Poverty: 20% of census tract families had income below poverty line in last 12 months.

[7] SEER Summ Stg 2000 = Surveillance, Epidemiology, and End Results (SEER) Program Summary Stage 2000.

[8] AJCC = American Joint Commission on Cancer.

[9] California-Sacramento includes Alpine, Amador, Calaveras, El Dorado, Nevada, Placer, Sacramento, San Joaquin, Sierra, Solano, Sutter, Yolo, and Yuba counties.

[10] Florida-Miami includes Broward, Hillsborough, Miami-Dade, Orange, and Palm Beach counties.